DOCUMENT RESUME

ED 023 301 56 EM 006 295

By-Stolurow, Lawrence M.; And Others

Pilot Studies of Principles of Programing. Comparative Studies of Principles for Programing Mathematics in Automated Instruction.

Illinois Univ., Urbana.
Spons Agency-Office of Education (DHEW), Washington, D.C. Educational Media Branch.
Report No-NDEA-7A-806; TR-9
Pub Date Jul 64
Note-64p.
EDRS Price MF-$0.50 HC-$3.30
Descriptors-Constructed Response, Educational Principles, Educational Research, Fixed Sequence, *Intermode Differences, *Mathematics Instruction, Mathematics Materials, *Programed Instruction, Program Evaluation, Program Length, Response Mode, *Time Factors (Learning)
Identifiers-UICSM, University of Illinois Committee on School Mathem

High school students took part in a series of pilot studies on the variables involved in step size and sequencing in programed instruction. Programed learning materials of the University of Illinois Committee on School Mathematics (UICSM) were varied as to covert and overt trial sequences, use of teachers, type of prior practice, order of frames, and size of step. Learning time was not significantly different when sequences of overt and covert trials were varied. When teachers lectured before or after students used programed materials, no significant difference in performance was found between the two conditions. Type of prior practice was varied three ways: stimulus-practice, mediator-practice, and response-practice, all of which were equal in their effect on performance. The order of appearance of items in a program was found to make no difference in learning. In a final study comparing large and small steps, the large step program was devised by omitting non-essential frames from the original program. A t-test indicated a significant saving of time for subjects using large step materials. It was concluded that studies which further investigate these variables should be based on larger samples. (LS)

UNIVERSITY OF ILLINOIS
Urbana, Illinois

Pilot Studies of Principles of Programing

Lawrence M. Stolurow
Ellen F. Rosen
Gerald L. Frincke
David Batchelor
Clark Himmel

COMPARATIVE STUDIES OF PRINCIPLES FOR PROGRAMMING MATHEMATICS IN AUTOMATED INSTRUCTION

Technical Report No. 9

July, 1964

Co-Investigators:

Lawrence M. Stolurow
Professor, Department of Psychology
Training Research Laboratory

Max Beberman
Professor, College of Education
University of Illinois Committee
on School Mathematics (UICSM)

Project Sponsor:

Educational Media Branch
U. S. Office of Education
Title VII

Project No. 711151.01

UNIVERSITY OF ILLINOIS
Urbana, Illinois

Pilot Studies of Principles of Programing

Lawrence M. Stolurow
Ellen F. Rosen
Gerald L. Frincke
David Batchelor
Clark Himmel

Technical Report No. 9

July, 1964

COMPARATIVE STUDIES OF PRINCIPLES FOR PROGRAMMING MATHEMATICS IN AUTOMATED INSTRUCTION

Co-Investigators:

Lawrence M. Stolurow
Professor, Department of Psychology
Training Research Laboratory

Max Beberman
Professor, College of Education
University of Illinois Committee
on School Mathematics (UICSM)

Project Sponsor:

Educational Media Branch
U. S. Office of Education
Title VII

Project No. 711151.01

TABLE OF CONTENTS

List of Tables

Figures Page

Chapter One[1]

Introduction

This report serves as a preliminary description, summary, and partial evaluation of some of the variables involved in step size and sequencing in programed instruction. Each variable is first introduced, then briefly described and finally the preliminary pilot study of the variable is described (Chapters Three-Seven). Chapter Three in particular presents some of the operating procedure for the preliminary work and some teacher comments. In Chapter Eight, a larger series of experiments is presented which evaluates, in more detail, the effects of those variables which were felt to be of potentially greater interest.

The interested reader is referred to Ellen F. Rosen and L. M. Stolurow, Description of variables and their implementation in studies of the principles of programing, Tech. Rep. No. 8. Urbana, Ill.: Univer. of Ill., Training Res. Lab., USOE Title VII, Proj. No. 711151.01, July, 1964, for a more complete description of the variables used in the extended study series. The report contains examples and descriptive data which make the variables much more comprehensible. A comprehensive summary of these studies also is included in the final semi-annual report (Quarterly reports 9 and 10, Urbana, Ill.: Univer. of Ill., Training Res. Lab., USOE Title VII Proj. No. 711151.01, December 6, 1962 - June 6, 1963).

---

[1]The authors wish to acknowledge the assistance and advice of John Gilpin and O. Robert Brown of UICSM.

Chapter Two

Overt and Covert Responding

Psychological learning theories differ in the importance which they place upon responses (see Spence, 1951). In meaningful verbal learning Kaess and Zeaman (1960) found a difference in error elimination rates which was related to whether or not the students made overt responses. They suggested that with overt response the connections are stronger and, therefore, errors are more difficult to eliminate than when responses are made covertly. Stolurow and Walker (1962) found that students performed as well on a test of achievement following programed learning regardless of whether they made overt responses or not. However, they also found that when students do not make overt responses they go through the program more rapidly. Consequently, the amount a student learns per minute of study time is greater when no overt response is required.

In the present study, these findings were considered in formulating an hypothesis about the relative effectiveness of two different sequences of practice trials. One sequence consisted in the use of covert response on the first trial and overt response on the second trial. The other sequence consisted in the reverse -- overt then covert. The hypothesis was that it would make a difference in performance on an achievement test given after two trials of practice if students used a sequence that required overt then covert response rather than the reverse sequence, provided it was more important to have weaker wrong than stronger correct responses established initially. In other words, if the elimination (or extinction) of the wrong responses was more difficult than the strengthening of correct

responses, then the covert-overt sequence would be superior to the overt-covert sequence. If this hypothesis is true, it also would follow that the advantage of the covert-overt sequence would increase as the base rate for the proportion of errors increased. Therefore, the covert-overt sequence would have a greater advantage over the overt-covert sequence if the program had a relatively high error rate.

One problem in testing this hypothesis is in identifying the range of error rates in programs over which the one sequence is superior to the other to a sufficient extent to be apparent in terms of the net effect on performance. Another problem is in determining the number of trials in the sequence over which the net result would be greatest. If, for example, a program were to have a base rate for error of 15 per cent and if it were assumed that the probability that an error would be extinguished in one trial was .6 whenever that error had been made with <u>covert</u> practice on the previous trial, then on the second trial of a covert-overt sequence the per cent correct responses would be 85 + (.6 x 15) = 94. On the other hand, if the extinction probability was only .4 whenever the error was made under <u>overt</u> practice, then the net effect after the second trial would be 85 + (.4 x 15) = 91. These net effects would be expected to be the same if the sequences had been covert-covert and overt-overt, respectively. However, had a three trial sequence been used, then the four conditions would not be equivalent any longer; however, they would fall in between the limits of the two pure sequences. The covert-covert-covert sequence on the third trial would add (.6 x 6) or 3.6 items to yield a total of 97.6 per cent correct. The overt-

overt-overt sequence, on the other hand, would net (.4 x 9) or 3.6 items on the third trial to yield 91 + 3.6 or 4.6 per cent correct. Thus, the added trial would increase both totals but not the difference between them. If a fourth trial were given the covert sequence would add only 1.44 items, but the overt sequence would add 2.16 items, thereby reducing the difference between the two pure sequences. Since these functions approach a limit at different rates,but from a high initial level, it would appear that the maximum difference in achievement will occur earlier in practice rather than later.

With respect to <u>time</u> as a dependent variable, the main savings would be expected in relation to the overt response condition. By having a covert trial precede an overt one, the overt performance time should be reduced. This would be expected from the fact that the covert trial would delimit the response repertoire for the student and this would reduce latency. Furthermore, specific responses might be consolidated during the covert trial so that the actual response times would be shorter. Both of these effects would be expected to summate and be determined by the correct responses more than by the incorrect ones in a learning task where the base error rate is low. Delays in response would be produced by the errors made, however, and each error, if corrected, could be expected to require; at a minimum, the insertion of one practice trial. Consistent with the previous assumption about the asymetrical effect of overt and covert practice, it would be expected that more than one practice trial would be required if the erroneous response had been overt. This would lead to the expectation that the

overt-covert sequence would lead to greater time on the overt trial than the covert-overt sequence would. Thus, there are several factors which would lead to the hypothesis that the covert-overt sequence would result in less time for the overt trial than would the alternative sequence.

In summary, if it is assumed that a covert response to a stimulus produces an S-R pairing which is more amenable to change than a response which is made overtly, then students who read and respond covertly to a program and then reread and respond overtly to the same program (covert-overt group) should show greater performance on an achievement test and require less time to complete their overt trial than students who first study the program overtly and then covertly.

Method

Materials

In order to test this hypothesis, Part 101-R of the UICSM programed learning materials was used with its corresponding part test to measure achievement. This part was chosen because the responses called for are generally short and, therefore, easily kept in immediate memory. The responses generally consist of one or two words per frame. Furthermore, the booklet is linear and in "Zebra" format. Consequently, all students would cover the same material. The topic covered by the booklet is the meaning of the concepts of number and numeral.

Procedure

The subjects were divided into two groups on the basis of a multiple regression combination of the STEP test (Mathematics 3A) and TOGA test scores. The students were assigned to groups so that the mean regressed scores for the two groups were nearly identical. The covert-overt group was given the booklet and no worksheet with the following instructions.

> "In a few minutes I will let you begin the first booklet of the series of sixteen similar booklets that you will be studying during this four week session. All sixteen of these booklets allow you to learn in an interesting way some of the materials and concepts that are contained in ninth grade algegra and mathematics."
>
> "Turn to page zero of booklet 101-R." [Then read page (0) to them. Then give these instructions:]
>
> "In a moment I want you to begin work on this booklet but do not use pen, pencil, or paper to write out or figure any of the answers to the questions. You won't have any worksheets or answer sheets, so no writing is necessary. After you have read a question carefully and studied it, hold your answer in your head and then turn the page to see if it was correct. Then go on to the next question, again reading and studying it carefully and hold your answer in your head, then turn to the next page to see the correct answer, and so on until you have finished the booklet. Do not skip back to review questions that you have already completed and do not skip around in the booklet answering questions because this will only confuse you. You should go through the booklet from one page to the next slowly enough so that you understand the questions and their answers well.
>
> As soon as you finish the booklet, close it on top of your desk. I will then collect your booklet and give you a written test which covers the material taught in the booklet that you just finished."

The overt-covert group was given the booklet with a worksheet with the following instructions:

> "In a few minutes I will let you begin the first booklet of the series of sixteen similar booklets that you will be studying

during this four week session. All sixteen of these booklets allow you to learn in an interesting way some of the materials and concepts that are contained in ninth grade algebra and mathematics.

Turn to page zero of booklet 101-R." [Then read the page (0) to them. Tell them to begin.]

As each student finished his booklet the first time he was given a part test (allowed 20 minutes). On the next day the covert-overt group was given the 101-R booklet with the worksheet and told to work through the book and write their answers in the indicated places; the overt-covert group was given the booklet and given the part of the instructions that they had not heard the previous day. As each student finished, he was readministered the part 101 test.

Results and Discussion

Test Scores

The results of test scores are presented in Table 1. None of the *t*'s which were calculated for test 1, test 2, or the comparison of test 1 and 2 were significant at the .05 level. However, there did seem to be some improvement in performance after the second reading even though it was not a statistically significant increase. The data suggest that with a larger sample size such differences might be statistically significant.

The critical finding, however, refutes the theoretical analysis on which the hypothesis was based. The overt-covert group showed larger gains than the covert-overt group (1.72 vs. 1.28, respectively). While neither of these exceeded the .05 level, it is true that the overt-covert group fell

between the .15 and .10 levels; whereas the covert-overt group fell between the .25 and .20 levels. Thus, while the data are not conclusive, the trend is opposite to the hypothesis.

Learning Time

Table 2 presents the data summary of the time spent on the booklet when done overtly. Since the overt run through was Group I's second time through the program, they should have spent less time. They, in fact, showed about a 15 per cent time savings. Although in the direction dictated by the above prediction the data are not significant. Here too, a larger sample size might lead to significant results.

Whereas the trend of the achievement data was contrary to expectation, the time data are not. However, there is more than one basis for the expectation that less time would be required for the overt response trial when it was after, rather than before, the covert trial.

The present treatment of the overt vs. covert response problem is based upon a fine grain analysis of behavior. It suggests that the methodology used in the study of this problem may be one in which probability estimates of changes from incorrect to correct responses could be obtained, and in which the time taken in going from frame to frame could be secured so that the time taken after an incorrect response could be related to that taken after a correct response. A computer based teaching system that recorded latencies as well as correctness of response could provide the required technology for a test of the proposed analysis.

Summary

In summary, the data collected are not definitive. While the trends in performance scores are not in the predicted direction, the time scores are. Another study using a larger sample size seems indicated in order to further test out this hypothesis and to provide estimates of the parameters that could be the determinants of the results, e.g., the probability of a wrong response being corrected, and the times spent on frames when a wrong response was made.

Table 1

Mean and Variance for Part 101-R Test Scores

Comparison	Test 1		Test 2	
	Mean	s^2	Mean	s^2
Group I (covert-overt)	16.29	13.57	17.57	13.95
Group II (overt-covert)	16.57	5.95	18.29	10.90

Table 2

Mean, Variance, and t for time (minutes) spent on 101-R Booklet when Done Overtly

	Mean[a]	Variance	t	df	p (one-tailed)
Group I (C-O)	93.71	594.57	1.45	12	.10 > p > .05
Group II (O-C)	109.86	274.48			

[a]This is approximately a 15 per cent savings.

Chapter Three

Teacher's Report[2]

UICSM-PIP 1963 Summer Institute

Physical Facilities

Three classrooms were available for the course: the first was used generally by all the students to work through the booklets; the second was used only when an experimental program made it necessary to split the class into two groups, and the third was used as a laboratory-type classroom where teaching machines and additional programs were available.

Course and Operating Procedure

The course was set up as a four-week class, meeting five days a week from 9 a.m. until 12 noon. There were two ten minute breaks each day at 9:50 and 10:50. Because some students expressed a desire to attend summer camp for a week, the class was extended for one week so that those students could make up the work that was missed.

Testing. The first day began with an explanation of the course, what would be done, and some of the administrative details that would be required during the first week. The class was then tested using Tests of General Ability (TOGA) grades 9-12 and Sequential Tests of Educational Progress (STEP), Mathematics Form 3A. These tests provided data concerning general

---

[2]We are grateful to Urbana School System personnel and, in particular, to Urbana Junior High School, Urbana, Illinois, which served as host for this institute. The four week programed instruction course began on June 17, 1963, and was taught by Dave Batchelor of the University of Illinois.

intelligence (measured by TOGA) and the level of mathematical aptitude (indicated by STEP). After the testing period was completed, the students were given a short questionnaire to be taken home and completed by their parents. This questionnaire ascertained whether there would be any objections if the students were asked either to come early or to stay late the days on which experimental booklets were used.

Programed learning. That morning and each morning thereafter the booklet on which the student was working was placed on his desk with the work sheet enclosed. Upon completing a booklet, he was instructed to place the worksheet inside the front cover of the booklet and return it to the teacher. The teacher would then hand out the part test (if there was one) with the time that it was due written at the top. If no test followed the booklet, the student was given the new booklet and worksheet.

Experiment I

The first experiment compared two sequences to determine their relative effectiveness as practice conditions. More specifically, it compared two practice conditions which were identical except for the sequence in which the two component conditions were used.

The students actually began to work through the program on Tuesday. The first booklet (101R) was an experimental booklet for which the class was divided into two sections so as to equate them in terms of their STEP and TOGA scores. The two groups were called the overt-covert group and the covert-overt group. Instructions were given to the two groups as indicated in Chapter Two.

The students then began working through the booklet. As each student finished the booklet he was given the part test. The remainder of the class period was spent working on the teaching machines.

The following day (Wednesday) the two groups were interchanged with the above instructions reversed for each group. Upon completion of the booklet and tests, the class was allowed to start working on booklet 102. Booklet 103 was begun immediately upon completion of 102. Since one student was to be gone the following week, he was given booklet 103 to finish at home Thursday evening. Friday morning he took the 102-103 part test and was given the lead-follow experiment (104) singly that day. All of the covert-overt group finished the booklets in 76 minutes or less while the overt-covert group took from 65 to 146 minutes. By noon on Friday all students had completed booklet 103 and taken the part test with the exception of one.

A new student joined the course on Monday of the second week but was omitted from both the covert-overt study and the lead-follow study. The students who had finished the part test on 102-103 were again divided into two groups and put into separate classrooms. One group -- the lead group -- began immediately working on booklet 104. The other group -- the follow group -- engaged in class discussion concerning the material contained in booklet 104, namely, multiplication of real numbers. The same type of'discovery teaching' was employed by the teacher in the discussion class as is used in the booklets. The teacher served only as a guide in developing and guiding the pupils toward the correct deductions. This discussion period lasted nearly 90 minutes. These students were then given the 104 booklets to work through. As soon as all the students in the lead group had finished their booklets, they were then given a similar discussion

period over the material just covered by them in the booklets. This discussion lasted only 40 minutes since they knew in what direction they were heading. During both sessions of the discussion, a tape recorder attempted to monitor the class but the quality was poor and no transcript could be obtained. It should also be noted here that the students enjoyed the discussion class very much and many of them indicated a preference for this type of class each day.

Following the discussion session, both sections -- the lead group at the completion of the lecture and the follow group at the completion of the booklets -- were given the same test over the material in booklet 104.

The following day (Tuesday) the students began work on booklet 105. Those who completed the booklet were given the part test in preparation for the S-M-R study (106). Those students who did not finish the 105 booklet in class were allowed to take the booklet home and finish it as homework. They were given the 105 part test the following morning.

Since the three versions of 106 were rather short in comparison with the other booklets, all of the students were able to finish it by noon on Wednesday. This includes both students who took the 105 part test on Tuesday and students who took the test on Wednesday morning. Although 106 was an experimental booklet, no change was necessitated from the normal procedure for booklets and part tests. Those students who finished the test over part 106 were given booklet 107.

Thursday and Friday the students continued to work through the booklets; the majority of the students had completed booklet 108 by Friday noon. Since these booklets were not experimental, those students who were behind

were given the booklet they were working on to take home on Thursday and Friday afternoons. Otherwise, there was no change in the normal procedure.

On Monday of the third week (July 1) it was announced by the Director of the Urbana Summer School that the students would be dismissed Friday, July 5, so that they could have a long Fourth of July weekend. This caused no problems for the course since the 110.5 booklet was not fully prepared until the following Monday. It was decided that all pupils would finish 110 by noon on Wednesday so that everyone (with the exception of those at summer camp) would begin booklet 110.5 on Monday. Thus, this week produced no variation from the normal procedure previously outlined.

On Monday, July 8, the experimental booklet 110.5 was given to the students. The only change in the normal procedure necessary was that the students were required to stay until the booklet and part test were finished.

Since the questionnairs mentioned earlier indicated n problems with early or late classes and since the class was running behind the four week schedule, it was indicated to the students that class would begin at 8:30 a.m. for the remainder of the course. It was also indicated that the students would be given booklets to work as homework. Thus, most of the students were given booklet 111 as homework for Monday afternoon and evening.

As soon as a student completed the part test over booklet 111, he was given the experimental 112 booklet if he had time to complete it in class. If not enough class time remained to complete 112, the student was given an English program to work as a supplementary activity for the rest of the morning.

Booklets 113, 114, and 114.5 required no change in the normal procedure. Immediately upon completion of booklet 114.5, the student was given the part test. As soon as the part test was handed in, the final Unit I test was given in the following manner: the student was instructed to read the test directions printed on the front cover of the test booklet at the teacher's desk. He was then given a plain sheet of paper and an answer sheet and told to circle the correct answer on the answer sheet and to use the plain paper for scratch work. He was instructed that he was to make no marks on the test booklet and that he would have thirty minutes to work on the test. Any questions the student had concerning directions were answered and the student was instructed to begin. The times were kept by the teacher.

The next school day, the students were given the STEP Mathematics 3B as a posttest. It was administered exactly as indicated in the STEP manual. Two students who were at camp for a week had to make up the material and were given the STEP test immediately following the Unit I final test.

Teacher Comments

(1) As indicated earlier in the paper, many of the students expressed a desire to have a discussion period in conjunction with the booklets (as with 104). Many of the students became disinterested toward the end of the course and seemed to be working through the booklets simply to complete them rather than to learn the material presented.

(2) It is my opinion that a four week session is not long enough to cover the material. Some of the slower students were rushed too much in order to finish the course in four weeks and some of the students had to return for a fifth week to finish. If the entire Unit I is taught (only 16 of the 18 booklets were used), I feel a minimum of six weeks should be allowed. This would allow not only for all students to finish the booklets but also give some time for class discussion. Of course, this would necessitate some sort of extra activity for the faster students who might finish early.

(3) If the course is taught in the summer, some arrangement should be made with the participating school to give at least partial credit to the students if at all possible. This would no doubt have a desirable effect on the student's motivation.

Chapter Four

Use of Programs with Teachers

Purpose

In addition to using the programed materials as the only mode of instruction, it is also possible to use them as either preparatory material to the teacher's lecture (lead) or as a supplement to the material the teacher presents (follow). The purpose of this study was to determine which of these two modes of presentation produces better performance. The lead group should perform better on a test over the material presented since they have a chance after becoming acquainted with the materials to ask questions and may be able to resolve any difficulties they have with the program. Part 104 was chosen to study this problem; the booklet and end test were not altered.

Method

The students were divided into two groups of six students each. The groups were formed on the basis of performance on a part test covering the material in Part 102 and 103. The groups were arranged so that their mean scores on the Part 102-103 test were equal.

Not all of the students heard the lecture appropriate to their group at the same time, but the follow group's lecture was the same for all sessions. The lead groups lecture was basically a question and answer session and all sessions were kept as similar as possible.

The follow group heard their lecture from the teacher first, after which they were given Part 104. When they completed the booklet they were given 20 minutes to complete the part test.

The lead group was given the Part 104 booklet, after which they were given their lecture. Upon completion of the lecture they were given 20 minutes to complete the part test.

Results and Discussion

The means and variances of the two groups on Part 102-103 test are presented in Table 1. Initially, the two groups do not differ significantly ($.400 > p > .35$).

Since the groups were comparable at the beginning of the study, a simple comparison of means is used for the analysis of the Part 104 tests. The data are presented in Table 2 along with an F-test for homogeneity of variance which was accepted ($.250 > p > .100$). The t-test led to the acceptance of the null hypothesis that there is no difference ($.400 > p > .350$). The difference is in the predicted direction; however, the lead group performed better than the follow group, whereas on Part 103 the reverse was true. This suggests that perhaps there is some slight support for the hypothesis.

The time spent on Part 104 is summarized in Table 3. Also presented in this table is a t-test of the difference between the means which were not significant. Nonetheless, the data are in the expected direction: one would expect the follow group to spend less time on the booklet than the lead group since they have previously covered the material with the teacher.

Table 1

Mean and Variance of Scores on Part 102-103 Test
and Comparison of the Means

Group	Mean	Variance	t	df	p (one tailed)
Lead	17.50	6.70	.366	10	
Follow	18.16	10.97			

Table 2

Data and Analyses of Part 104 Scores

Group	Mean	Variance	F	df	p	t	df	p
Lead	14.33	61.47	3.16	5,5	.250 > p > .100	0.273	10	.40 > p > .35
Follow	13.33	19.47						

Table 3

Summary of Time (Minutes) on Part 104

Group	Mean	Variance	t	df	p (one tailed)
Lead	83.5	357.1	0.54	10	$p \tilde{=} .300$
Follow	78	266.8			

Chapter Five

Type of Prior Practice

Purpose

The objective of this study was to find out whether having stimulus practice, mediator practice, or response practice before learning a task of which these were components produced a savings in learning the task. In order to implement this study, Part 106 of the UICSM programed materials was broken up into three smaller booklets: Part 106 (stimulus practice), Part 106 (mediator practice), and Part 106 (response practice). The internal quiz section of Part 106 was made into an end test. The breakdown of these materials is given in Table 1.

Method

Tasks

The stimulus booklet consisted of items which asked the student to identify two identical stimuli, i.e., to punctuate expressions so that they became identical. The mediator booklet introduced the student to the notion of binary operation and to the left-hand convention for simplifying when binary operators are present. The student was asked to apply these by finding binary operators and to punctuate expressions according to the convention. The response booklet taught the student how to simplify expressions rotely and asked the student to simplify expressions. In both the stimulus and response booklet the didactic conceptual device of the UICSM computer was introduced, but it was not introduced in the mediator booklet although the student learned how to prepare topics for it.

Since Parts 106 and 107 are so similar, it was decided to leave Part 107 untouched and to use the Part 107 test as it is. Part 107 presented the task to be learned and 106 familiarized the student with the elements used in the task.

Students

The students were divided into three groups on the basis of their prior performance on the UICSM Part 105 test. Each group was chosen so that their mean performance on Part 105 test was nearly identical. The number of students in each group and their scores on Part 105 are given in Table 2 and in Figure 1.

Each of the three groups was assigned to a treatment condition and was given the appropriate booklet. As each student completed his booklet, his time was recorded and he was given the Part 106 test for which he was allowed 20 minutes. When he completed the test he was given Part 107. Upon completion of 107 (for which time scores were also recorded), the student was allowed 20 minutes to complete the Part 106-107 test.

Results and Discussion

Initial performance. The summary of the analysis of variance of the mean scores of the groups on Part 105 is presented in Table 3. The analysis ($F < 1$) shows that the groups were indeed equal in performance on Part 105. The groups were thus indeed equal in performance prior to their beginning Part 106.

Learning. Table 4 presents the analysis of variance on the mean scores on Test 106 (which are given in Table 2 and Figure 2). The scores for the groups are not significantly different ($F < 1$).

Table 5 is a summary of the analysis of the mean scores on Test 107 (which are presented in Table 3 and Figure 3). There are no significant differences between the three groups in performance on this test.

The conclusion to be drawn is that none of the three treatments led to superior performance on the task. This same conclusion also is indicated in terms of time spent on the booklet. The analyses of the time spent on Parts 106 and 107 are presented in Tables 6 and 7, respectively. Thus, the hypothesis that prior experiences with different components of a subsequently learned task will affect performance on the task seems to be rejected.

Table 1

Breakdown of Parts $\overline{106}$, $\underline{106}$, $\underline{\underline{106}}$

Part	No. of Pages	No. of responses required of the student
106 (m)	26	73
106 (s)	24	73
106 (R)	25	73
Test 106	2	33

Table 2

Number of Students and Mean Score of Groups on Part 105, 106 and 107

Part	Group	N	Mean Score
105	stimulus	5	7.00
	mediator	4	8.25
	response	5	7.40
106	stimulus	5	18.00
	mediator	4	15.50
	response	5	17.60
107	stimulus	5	4.20
	mediator	4	3.25
	response	5	3.00

Table 3

Summary of Analysis of Variance of Part 105 Scores

Source	df	MS	F
Treatments	2	1.975	< 1
Error	11	16.359	
Total	13		

Table 4

Summary of Analysis of Variance of Scores on Test 106

Source	df	MS	F
Treatments	2	7.755	< 1
Error	11	36.75	
Total	13		

Table 5

Analysis of Variance of Scores on Part 107

Source	df	MS	F
Treatments	2	1.98	< 1
Error	11	20.32	
Total	13		

Table 6

Summary of Analysis of Variance of Time on Part 106

Source	df	MS	F
Treatments	2	.54	< 1
Error	11	529.12	
Total	13		

Table 7

Summary of Analysis of Time Spent on Part 107

Source	df	MS	F
Treatments	2	6478.23	< 1
Error	11	8664.98	
Total	13		

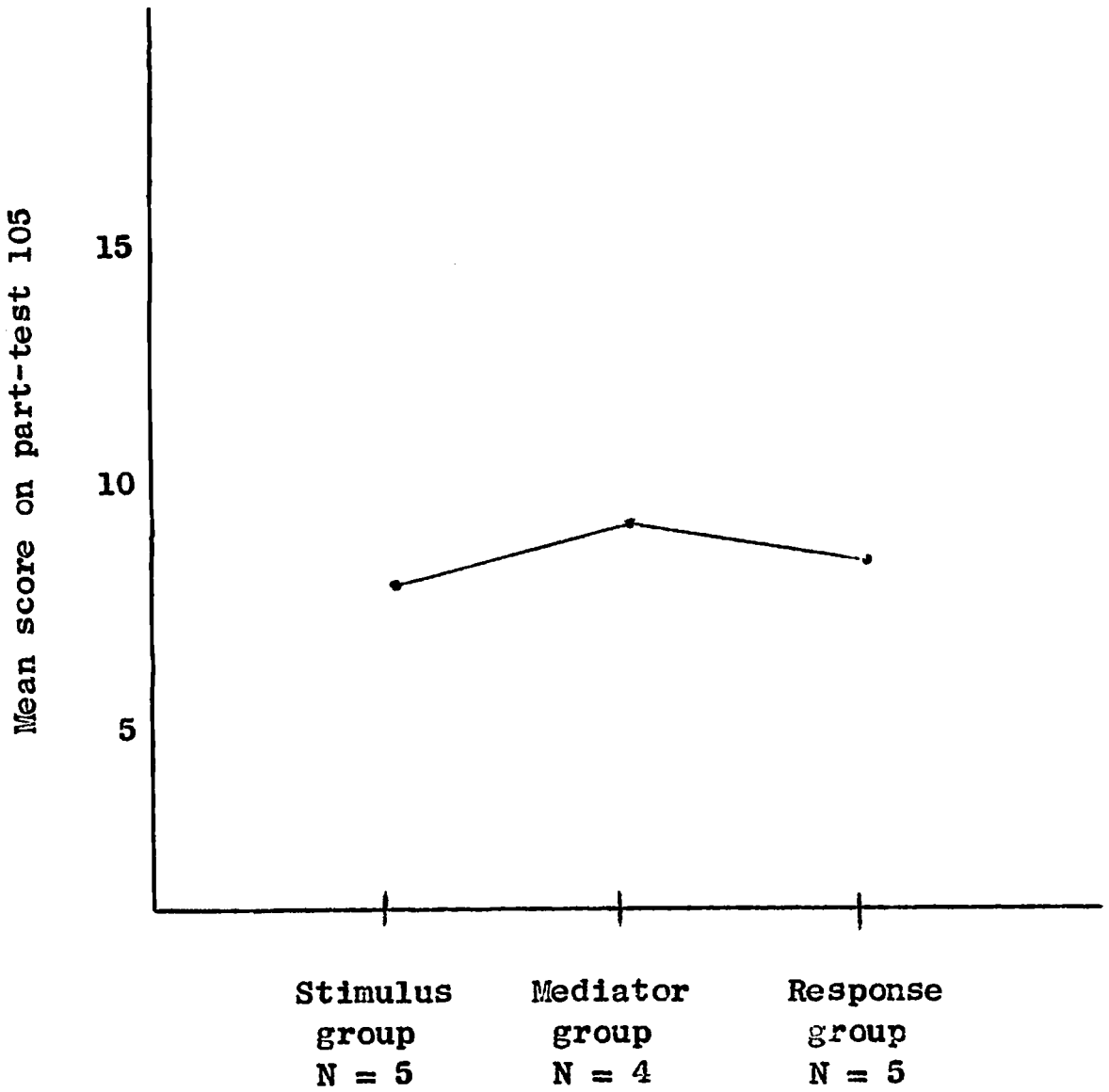


Fig. 1. Mean Score on Part 105 Test

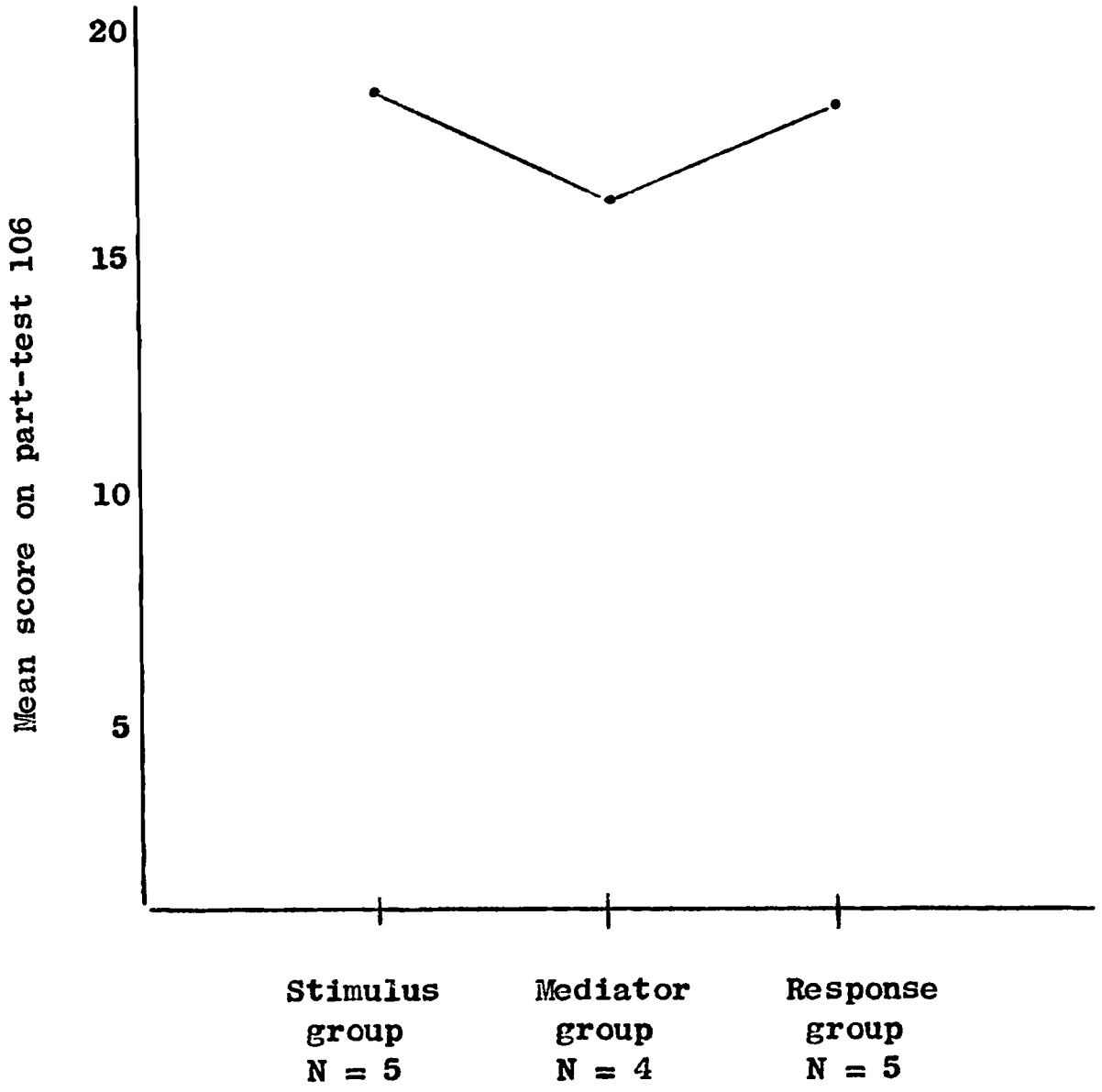


Fig. 2. Mean Score on Part 106 Test

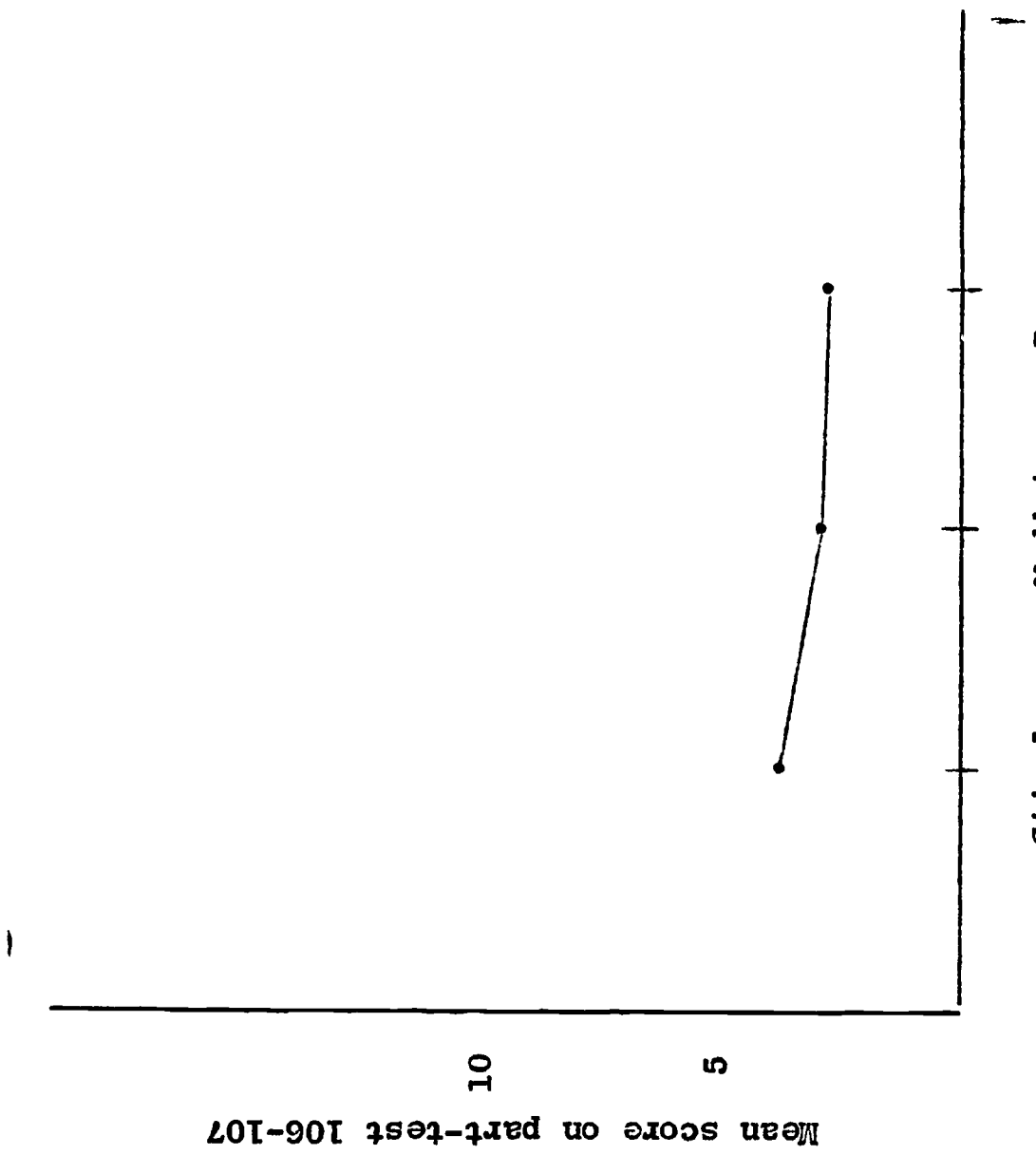


Fig. 3. Mean Score on Part 107 Test

Chapter Six

The Order of Frames

Purpose

The order in which items appear in a program should make a difference in learning. The items should be in a sequence which produced optimal effects (see Smith, 1962). In order to test the effect of sequence, one booklet (Part 110.5) was prepared in two forms. One form was unaltered as used in the 1962-63 studies; the second form was produced from the original by rearranging frames in essentially random sequence.

Hypothesis

The hypothesis is as follows: if there exists a sequence which is most efficient for the learner, a sequence also should exist which is least efficient. It seems intuitively obvious that a random sequence would be the least efficient. The effect of sequence would seem to be revealed most clearly when the material (content) is held constant while different sequences of frames are compared with one another.

By this hypothesis, the random sequence group ought to demonstrate poorer performance on an instrument which measures their grasp of the task learned since the random group must structure and organize the material into a logical sequence as well as learn the task. Thus, the random group should have more time to learn the task.

Method

Two forms were produced (Part 110.5 and Part 110.5) from the original form of Part 110.5. Part 110.5 was left unaltered but Part 110.5 was produced by using the frames of Part 110.5 and reordering them into an essentially random sequence; then a test was written by the teacher to cover this material.

Subjects. The subjects were divided into two groups on the basis of scores on a test covering materials from Parts 108-110. These materials were presented in such a manner that the students' mean performances on Part 108-110 test were equal. There were seven students in each group. As each student finished the booklet, he was given twenty minutes to complete the Part 110.5 test. Records were kept of the scores and of the amount of time each student spent on the booklet.

Results and Discussion

Table 1 contains a summary of the score data collected on both part tests. The t-test for the two groups showed that they were not different at the beginning of the experiment ($t = .30$ for 12 df) as revealed by scores on the test covering Part 108-110 which immediately preceded this instructional sequence. The scores obtained on the Part 110.5 test for the two groups indicated that both groups learned the task equally well ($t = .30$ for 12 df).

A summary of the time data is given in Table 2. Although the difference between the two groups is statistically nonsignificant, it is in the opposite direction from that which was predicted since the random group spent less

time on the program. Thus, the data do not support the hypothesis. However, at onset the random group (although statistically equal to the logical group) was slightly inferior in performance, and at the end of the experiment the random group was equal in performance to the logical group; therefore, the deduction concerning the amount of time spent on the booklet should hold true. The data demonstrated exactly the reverse phenomenon: the random group spent less time on the booklet. Another study using a larger sample needs to be run on this variable to find out if this time difference is a stable phenomenon.

Table 1

Mean, Variance of Scores on Tests

Part test	Group	Mean	Variance
108-110	Random	7.57	38.28
	Logical	8.50	29.08
110.5	Random	18.57	22.95
	Logical	18.57	24.28

Table 2

Mean, Variance and Comparison of Times Spent on Booklet 110.5

Group	Mean (minutes)	Variance	t	df	p
Random	95.86	182.48	-1.441	12	
Logical	148.29	9085.57			$.900 < p < .950$

Chapter Seven

Size of Step: An Exploratory Study of Some Measures

Purpose

There is a rather general agreement that instructional programs differ in the "size of step" they require of the learner. Unfortunately, there have been no a priori measures of step size available to the programer to permit him to pace his materials according to the learner needs.

This pilot study attempted to determine the validity and usefulness of some a priori measures of step size. In it, the approach was to identify and relate the members to a set of a priori measures which also were related to a posteriori measures of step size.

Method

The utility of the proposed a priori measures was determined in two ways: first, by comparing the a priori scale values obtained for the "large" step version of a program with those obtained for the "small" step version and second, by relating the performance of students on an end test -- those who used the "large" step version with those who used the "small" step version.

The validity of the measures was determined in two ways. The first was a homogeneity test obtained by intercorrelating the operationally different a priori measures. The second was a predictive test obtained by securing the correlations between the a priori measures and performance on the frames of the program.

A Priori Measures

The a priori measures of step size were the judgements of people familiar with programed instruction. The judges were asked to rate each step on a scale (-5 to +5) in each of four categories. An explanation of the four categories used is given in Appendix A -- Instructions to the Judges. The category measures were combined into an unweighted mean that was the a priori measure of step size used to represent each frame of a program.

Judgements

To obtain the a priori measures, two sets of materials were prepared for the judges. One consisted of the first section of Part 112 and the remainder of the second section of Part 112L. The other set of materials consisted of the reverse -- the first section of Part 112L and the second of Part 112. Each judge was given a copy of the instructions (Appendix A) and a rating sheet.

Materials

Two versions of UICSM programed mathematics booklet 112 were prepared for students and judges. One version (small step) was the original booklet 112. A second version (large step), booklet 112L, was developed by omitting frames from the original booklet, although no frame was omitted that introduced a new concept. Thus, step size was lengthened through the omission of frames.

Test

A Part 112 end-test was used to determine what the students learned. This test contained 30 items and required responses.

Students

Each of the two booklets was given to seven students at Urbana Junior High School. The two groups of students were equated for initial ability on the basis of a test covering Part 111 which was completed immediately before they began Part 112. As soon as the students completed the frames in Part 112, they took the Part 112 end-test.

Results and Discussion

A Priori Measurements of Step Size

The mean ratings for the categories are summarized in Table 1. There were five judges for each booklet. An attempt was made to use a technique for estimating judge reliability by using the analysis of variance (single factor-repeated measures) presented by Winer (1963, p. 127), but some of the assumptions were excessively violated. The coefficient of variation was computed for the ratings and is also presented in Table 1. It indicates the percentage of the mean for the standard deviation.

A Posteriori Measurements of Step Size

The proportion of correct responses (all items on a page correct) was computed for both groups. The average proportion of students correct per page for both booklets, the standard deviations, and the coefficients of variation are presented in Table 2.

Interrelations of A Priori and A Posteriori Step Size

The two measures of step size were intercorrelated: the total (sum of the category) rating score was correlated with the proportion of correct responses on the questions immediately subsequent to each step. For example, the judgement of the step from page 1 to page 2 was correlated with the number of correct responses to the questions asked on page 2. These intercorrelations are presented in Table 3.

It appears that although the categories intercorrelate highly among themselves, they do not correlate highly with the a posteriori measure. This could be due to the small number of students available.

Initial Performance

The means and variances of the two groups on tests for Part 111 and 112 are presented in Table 4. Clearly, at the outset of the experiment, the two groups did not differ significantly in original ability as measured by the Part 111 end-test ($t = .23$, $df = 12$).

Final Performance

The Part 112 end-test scores were in the direction opposite of that predicted, the long step group was superior although the probability statement for this is not at the usual significance level ($.850 > p > .800$). Furthermore, this represents a shift from Part test 111 where the small step group was slightly (though not significantly) superior to the large step group in achievement.

Summary

The summary of the data for the time the students spent on the booklets is presented in Table 5. The mean time spent on the large step booklet was indeed significantly shorter ($.05 > p > .025$); the mean saving of time was about 40 minutes. For the reader who is interested in the relative composition of the two booklets, and as an aid to understanding the difference in time, Table 6 presents the number of pages in each booklet and the number of required responses the student had to make.

Table 1

A Priori Step Size: Means, Standard Deviations (S.D.) and Coefficients of Variation (V)

Size of step	Statistic	Category				
		Concept	Vehicle	Numeral	Response	Total
Large step (112L)	Mean	.105	-.024	.004	.007	.019
	S.D.	.529	.865	.794	.725	.549
	V	503.81	3604.17	1950.00	10357.14	2889.47
Small step (112)	Mean	-.085	.011	.010	-.004	-.017
	S.D.	.646	.593	.654	.668	.401
	V	760.00	5390.91	6540.00	16700.00	2358.82

Table 2

A Posteriori Step Size: Means, Standard Deviations (S.D.) and Coefficients of Variation (V)

Proportion correct	Size of step	
	Large (112L)	Small (112S)
Mean	.655	.487
S.D.	.209	.278
V	31.91	57.08

Table 3

Intercorrelations of A Priori and A Posteriori Measures of Step Size[a]

Category	Category					
	Concept	Vehicle	Numeral	Response	Total	Proportion correct
Concept		.544**	.443*	.419*	.754**	.055
Vehicle	.222		.248	.399*	.746**	.016
Numeral	.070	.236		.522**	.738**	-.135
Response	.442	.208	-.026		.777**	.111
Total	.697**	.641**	.511**	.659**		-.018
Proportion correct	-.081	.046	-.092	-.298*	-.177	

[a]Small step size (112) given below main diagonal; large step size (112L) given above main diagonal.

* $|r| \geq .288$, significant at .05 level, df = 45, two-tailed.

or

$|r| \geq .349$ significant at .05 level, df = 30, two-tailed.

** $|r| \geq .372$, significant at .01 level, df = 45, two-tailed.

or

$|r| \geq .449$, significant at .01 level, df = 30, two-tailed.

Table 4

Mean and Variance of Scores (Number Correct)
on Part Tests 111 and 112 for the Two Groups

Group	Scores			
	Part Test 111		Part Test 112	
	Mean[a]	Variance	Mean[a]	Variance
Small step (112)	12.43	17.95	9.00	6.33
Large step (112L)	11.86	24.14	10.57	14.95

[a] *t*-tests not significant.

Table 5

Summary of Time Data Collected on Part 112 and 112L*

Group	Mean	Variance	Standard error of the mean	t	df	P
Small step (112)	111.29	2078.57				
			21.629	1.902	12	.05 > p > .025
Large step (112L)	70.14	1196.48				

*The absolute difference between the two means could be misinterpreted since the average time per page for the short and long step programs was about the same; i.e., 2.1 and 2.00, for small and large step programs, respectively. The absolute difference reveals the differences in the number of steps.

Table 6

Composition of Booklets 112 and 112L

Booklet	Number of		
	Booklet pages	Responses	Worksheet pages
112	53	282	13
112L	35	179	9

APPENDIX A[2]

Instructions for Judges

We are interested in the similarities and differences in pairs of adjacent pages or "learning steps" contained in the accompanying booklet of programed instruction, and we would like your help in finding out how much these adjacent pages are different from and similar to each other with regard to the complexity (abstractness) of certain given characteristics of the material present in the pages. (The pages to be judged will be considered in serial order, i.e., pages 1 and 2 will be compared, then pages 2 and 3, then pages 3 and 4, etc. through the final two pages in booklet.)

We want you to rate the changes in complexity (abstractness) of certain characteristics in going from the first page of the pair to the second page on a scale from -5 through +5, with a rating of zero (0) representing no change in the complexity of a characteristic, ratings above zero representing progressively increasing complexity from the first to the second page, and ratings below zero representing progressively decreasing complexity from the first to the second page, so that a rating of +5 represents the most expreme change in complexity of a characteristic in either direction. If a characteristic is not present on either of the pages of the pair, record a zero (0) as your rating.

---

[2]Prepared and developed by Clark Himmel to conform to the dimensional requirements developed in work with a program or fractions by L. M. Stolurow with the assistance of Gaila Grubb.

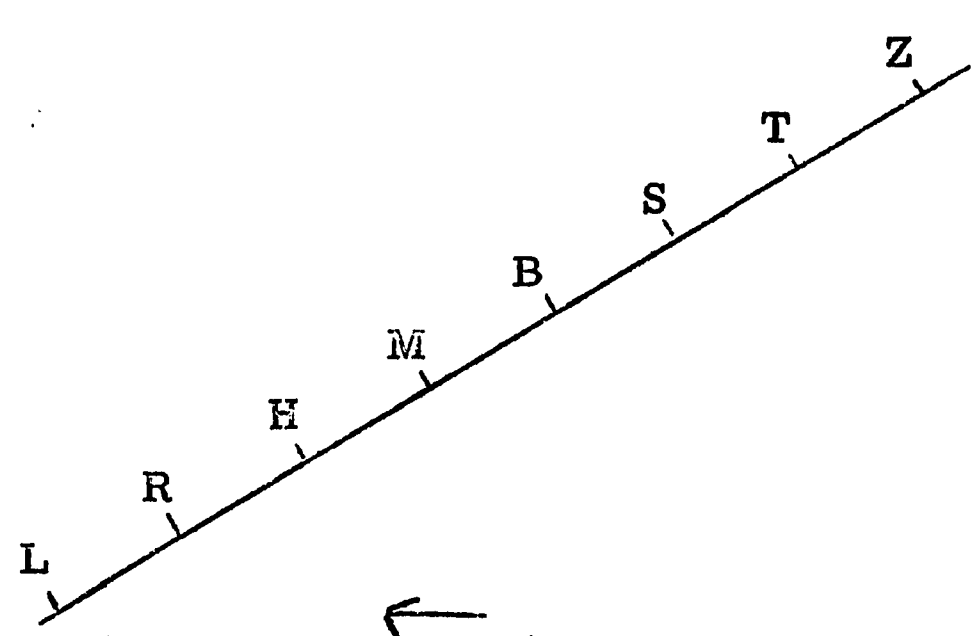
The four characteristics that we want you to consider are (A) the Concept, (B) the Vehicle, (C) the Numeral, and (D) the Response. A description of each of these characteristics, along with an example, and a rating guide is given below.

Concept: refers to the mathematical rule, principle, idea, or closely related group of rules, concepts, conventions, ideas, or principles in mathematics; such as, the associative principle of addition, or the axiomatic system in Euclidean geometry, or the idea of negative numbers.

You should be looking for one of the following: Changes in the complexity, in levels of description or in manner of presentation. You are to identify and rate these changes when leaving one concept and turning to another as they happen within two adjacent pages. Also, note changes in overall complexity when two or more concepts (or, if you prefer, "subconcepts") are presented simultaneously on one or both of the pair of pages being considered. For example, if only addition is presented on one page and both addition and multiplication are presented on the following page, the change probably is an increase in the complexity of this characteristic. If this occurred then the rating assigned to the pair of pages might be a +2 for the concept.

Vehicle: that which is used to help communicate or convey the concept (and the associated material) being presented by giving a concrete or exemplar background or "real setting" to the problems and expository material; such as, two airplanes traveling toward each other in a rate of travel problem in algebra, or the ledger entries for a retail business in a bookkeeping problem.

This characteristic is one which may not be present on all program steps. Consider the vehicle "a road with mile markers" for presenting the idea of real numbers (both positive and negative), where a trip from R to B (represented $\xrightarrow{\quad 3 \quad}$) is a +3



and a trip from T to B (represented $\xleftarrow{\quad 2 \quad}$) is a -2. If this same vehicle with no additions or deletions is present on both pages of a pair, the rating assigned would be zero (0). If it is absent only on the second page of the pair, the rating assigned would be

+5. (The above assumes that no new vehicle characteristics were introduced on either of the pages in the pair.) If something (diagrams, notation, verbal explanation) is added to the vehicle or a new vehicle is introduced in going from the first page to the second, a rating commensurate with the accompanying change in complexity should be assigned. If the same material were deleted from the second page, a rating commensurate with this change should be assigned.

Numeral: refers simply to all symbols for or representations of numbers presented, by the Roman numerals, Hindu-Arabic numerals, or others, plus their accompanying "operators" and "designators," such as +, ÷, $\sqrt{12}$, =, or -7, so that an entire expression like (+16 ÷ -4) x +2 = -8 would be considered under this characteristic.

Consideration should be given to changes in complexity in the types of numerals given on the pages. This should be relatively straightforward, since numerals and their "operators" and "designators" are presented in an explicit notation system. For example, a first page might present addition of simple three digit numerals while the next page calls for multiplication of the square roots of similar three digit numerals. Then the pair would probably receive a fairly high positive rating, perhaps a +3.

Response: refers to the particular answer(s) to be chosen, constructed or written, or in some way indicated by the student as he finishes the problem(s) or question(s) on a page.

Response complexity will vary due to the characteristics of the actual response given and due to the abstractness or difficulty of the specific question(s) or explicitly stated problem(s) to be answered or solved. For example, a response that would be relatively complex in the UICSM Unit I material would be one which is constructed or written by the student; for example, "the associative principle of addition." A relatively less complex response would be choosing one of two alternatives. The second facet of "response" to be considered is the nature of the problem(s) or question(s) to be answered. It also can be scaled in terms of complexity or abstractness. A question like "2 + 2 = ?" is probably less complex than a long and tedious word problem which also requires only a single digit answer.

Each of the characteristics on the pair of steps (pages) to be compared should be rated with regard to the change in complexity (or abstractness in the sense of being abstruse, more difficult to comprehend, ideationally complex or intricate) in going from one step to the next one.

On your rating sheets you will find the four characteristics listed as headings of four columns. Each pair of pages to be compared and then rated is listed at the left. When comparing pairs of pages, do not include the answers and "feedback" material (usually included between the statements "check your answers" and "record your results") in your considerations for rating. We are interested in having you rate the "instructional" and "question" portions of the pages.

Remember:

1. Rate Changes on the scale from

+5	Mid-point 0	-5
Increased Complexity	(no change)	Decreased Complexity

2. Consider the four following characteristics when rating each pair of pages:

   A. Concept

   B. Vehicle

   C. Numeral

   D. Response

3. For each characteristic consider the amount of change in your perception.

References

Educational Testing Service. Sequential Tests of Educational Progress (STEP). Princeton, N. J.: Educational Testing Service, 1957.

Flanagan, J. C. Tests of General Ability (TOGA). Chicago, Ill.: Science Research Associates, 1959.

Kaess, W. and Zeaman, D. Positive and negative knowledge of results on a Pressey-type punchboard. J. exp. Psychol., 1960, 60, 12-17.

Smith, Leone M. Programed learning in elementary school: An experimental study of relationships between mental abilities and performance. Urbana, Ill.: Univer. of Ill., Training Res. Lab., USOE Title VII, Proj. No. 711151.01, Tech. Rep. No. 2, August, 1962.

Spence, K. W. Theoretical interpretations of learning. In S. S. Stevens (Ed.) Handbook of experimental psychology. New York: Wiley, 1951, Pp. 690-729.

Walker, C. C. and Stolurow, L. M. A comparison of overt and covert response in programed learning. J. educ. Res., 1962, 55 (9), 421-429.

Winer, B. J. Statistical principles in experimental design. New York: McGraw-Hill, 1962.